

## Importance of the “What,” “When,” and “Where” of Mosquito Collection Events

DESMOND H. FOLEY,<sup>1</sup> RICHARD C. WILKERSON, AND LEOPOLDO M. RUEDA

Division of Entomology, Walter Reed Army Institute of Research, 503 Robert Grant Ave., Silver Spring, MD 20910

*J. Med. Entomol.* 46(4): 717–722 (2009)

**ABSTRACT** There is increasing need to apply established standards for recording data on mosquito collection events, because of the diversity of potential data providers, and the growth and interoperability of online databases designed to host these collection records. In particular, adequate taxonomic and georeference data are needed for geodatabases such as Mosquitomap (<http://www.mosquitomap.org/>) that map and compare these collection points with other spatial information in a geographical information system (GIS) setting. Accurately georeferenced collection data are crucial for understanding mosquito biogeography, ecology, and the impact of environmental changes, as well as for species distribution modeling, planning mosquito surveys, and for determining disease risk. We sampled representative published reports of new mosquito species records from 1980 in North America to the present to ascertain the quality of georeference information. Our results show that authors have increased the frequency of reporting georeferences but that they vary in the precision of the georeference, and some information, such as the source, date, and datum of the georeference, are usually not given. We discuss recently established standards for recording collection events, some relevant online resources available to researchers to assist them in their georeferencing, and the data input schema developed for the Mosquitomap database. We propose that the mosquito research community adopt data standards for recording and reporting the results of mosquito collection events to increase the value of these data. In particular, we recommend authors lodge voucher specimens and use a GPS set to the WGS84 datum.

**KEY WORDS** distribution, Mosquitomap, georeferencing, GIS, database

Collection location data are an underused resource for understanding mosquito biogeography, ecology, and the impact of environmental changes, as well as for species distribution modeling, planning mosquito surveys, and for determining disease risk. Accurately georeferenced mosquito collection data can be matched to remote sensing data of an appropriate resolution to answer many questions about the environmental determinants of mosquito distribution.

At a minimum, a mosquito collection event typically has information about what was collected, when the collection was made, and where the collection took place. The quality of this information varies and can be thought of as laying along a taxonomy, time, and space continuum. Low-quality information on all three dimensions would place the collection record in one corner of the graph, whereas high-quality information would place it in the opposite corner. Generally, re-

cording when a collection was made is less problematic than recording the “what” and “where” information. Improvements in our understanding of mosquito species inventory, and the availability of identification keys, descriptions, and molecular tools suggest that the potential quality of mosquito identifications has improved. Online resources such as the Systematic Catalog of the Culicidae (<http://www.mosquitocatalog.org/>) provide mosquito workers with up-to-date taxonomic information and online interactive identification keys. However, a survey of 80 recent ecological papers (Bortolus 2008) found that a majority did not have any taxonomic information to support species identifications, and only 2.5% reported that a voucher was kept. In this paper, we investigate the quality of the identification and location data of field collected mosquito species reported in mosquito publications.

There is an increasing need to apply established standards for recording data on mosquito collection events, because of the diversity of potential data providers and the growth and interoperability of online databases designed to host these collection records. In particular, adequate taxonomic and georeference data are needed for geodatabases such as the Global Biodiversity Information Facility (GBIF; <http://www.gbif.org>).

This research was performed under a Memorandum of Understanding between the Walter Reed Army Institute of Research and the Smithsonian Institution, with institutional support provided by both organizations. The published material reflects the views of the authors and should not be construed to represent those of the Department of the Army or the Department of Defense.

<sup>1</sup> Corresponding author, e-mail: [foleydes@si.edu](mailto:foleydes@si.edu).

<b>Report Documentation Page</b>			<i>Form Approved OMB No. 0704-0188</i>	
<p>Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p>				
1. REPORT DATE <b>2009</b>	2. REPORT TYPE	3. DATES COVERED <b>00-00-2009 to 00-00-2009</b>		
<b>4. TITLE AND SUBTITLE</b> <b>Importance of the 'What,' 'When,' and 'Where' of Mosquito Collection Events</b>			5a. CONTRACT NUMBER	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER	
<b>6. AUTHOR(S)</b>			5d. PROJECT NUMBER	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> <b>Walter Reed Army Institute of Research, Division of Entomology, 503 Robert Grant Ave, Silver Spring, MD, 20910</b>			8. PERFORMING ORGANIZATION REPORT NUMBER	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>			10. SPONSOR/MONITOR'S ACRONYM(S)	
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> <b>Approved for public release; distribution unlimited</b>				
<b>13. SUPPLEMENTARY NOTES</b>				
<b>14. ABSTRACT</b> <b>see report</b>				
<b>15. SUBJECT TERMS</b>				
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b> <b>Same as Report (SAR)</b>	<b>18. NUMBER OF PAGES</b> <b>6</b>
<b>a. REPORT</b> <b>unclassified</b>	<b>b. ABSTRACT</b> <b>unclassified</b>	<b>c. THIS PAGE</b> <b>unclassified</b>		
<b>19a. NAME OF RESPONSIBLE PERSON</b>				

org/) and Mosquitomap (<http://www.mosquitomap.org/>) that map and compare these collection points to other spatial information, in a geographical information system setting.

In the past, location information was usually limited to recording the country, province, or nearest town. For example, according to the Catalog of the Culicidae, the type locality for *Aedes cantans* (Meigen) is Europe, for *Anopheles algeriensis* Theobald is Algeria, and for *An. gambiae* Giles is the Gambia Valley, West Africa. Such geographical vagaries hamper taxonomic studies that seek to resample species from the original source. Later, longitude and latitude were sometimes provided from maps or paper gazetteers, usually after field collections were completed. However, such a posteriori georeferencing can inadvertently increase the risk of introducing errors. Another improvement was the development and use of standardized mosquito collection forms that included location fields, such as those promoted by Belkin et al. (1965) and later modified for the Mosquito Information Management Project (Faran et al. 1984). Early attempts to use computers to map mosquito collections (White and Grodhaus 1972, Faran et al. 1984), no doubt did much to promote more accurate georeferencing of mosquito collection events. The development of online gazetteers during the late 1990s also increased the ease of a posteriori georeferencing, and entomologists now have a wide choice of online gazetteers, such as Global Gazetteer V2.1 (<http://www.fallingrain.com/world/>), GEOnet names server (<http://earth-info.nga.mil/gns/html/index.html>), and Biogeomancer 1.2.1. (<http://www.biogeomancer.org/index.html>).

Perhaps the most important development for georeferencing was the widespread availability in the early 1990s of hand-held units accessing the Global Positioning System (GPS). Before May 2000, most GPS units used by civilians were subject to "Selective Availability," a signal degradation technique that limits accuracy to around 100 m or worse (McElroy et al. 1998). Most hand-held GPS units now promise errors of <10 m in open areas when using four or more satellites. Increasing access to the internet has also increased the georeferencing role of resources such as Google Earth (<http://earth.google.com/>) and Biogeomancer that allow the user to locate actual collecting sites. In addition, Biogeomancer automatically calculates the spatial error of georeferences derived from location descriptions, including headings and offsets, according to the point radius method (Wieczorek et al. 2004).

Guidelines for reporting georeference information for collection records have been developed (Chapman and Wieczorek 2006, MaNIS/HerpNet/ORNIS: <http://manisnet.org/search.shtml>). The ability for geodatabases such as GBIF and Mosquitomap to "talk" to one another has been assisted by the development of data standards that ensure interoperability. Data schemas such as Darwin Core (<http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome>) and ABCD (<http://wiki.tdwg.org/twiki/bin/view/ABCD/WebHome>), and the Distributed Generic Information Retrieval (Di-

GIR) protocol have allowed the deployment of a distributed biodiversity database network. The increasing use of mobile computing for field mosquito data capture suggests that it is timely that minimum standards for reporting be adopted. The recent development of ontologies using controlled vocabularies, such as the Mosquito Insecticide Resistance Ontology and Mosquito Gross Anatomy Ontology, available in a searchable format in VectorBase (<http://www.vectorbase.org/Search/CVSearch/>), offer new opportunities for standardizing terminology and for automated data exchange. These sorts of initiatives have been largely driven by advances in genomics and information technology rather than mosquito studies, and a challenge is to maximize the involvement of the mosquito research community in their development.

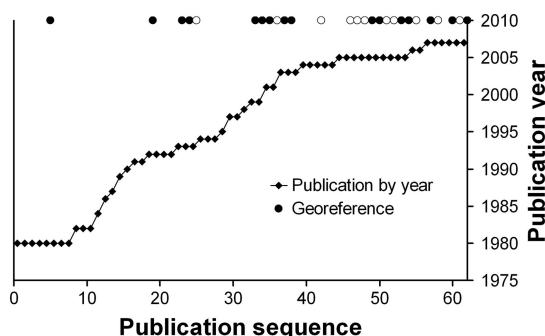
With so many developments to assist geodatabasing collection records, we were interested in finding out whether mosquito workers are both cognizant of the importance of georeferencing and whether they are taking advantage of these advances. We surveyed the georeferencing and taxonomic information within one type of mosquito collection report, i.e., publications in representative refereed journals from North America of the first occurrence of a species for an area. Because these observations are important for understanding invasive species and the threat of emerging diseases, the results were seen as a conservative estimate of the state of georeferencing of mosquito collections more widely.

## Materials and Methods

We sampled the literature for records of the first occurrence of a species for an area. We used the list of references in Darsie and Ward (2000) for the period 1981–1999 as a basis and largely confined later observations to published reports for the United States and Canada appearing in the *Journal of the American Mosquito Control Association* and the *Journal of Vector Ecology*. As such, this survey was meant to be illustrative rather than comprehensive.

## Results

A total of 62 references were obtained. Figure 1 shows that the frequency of georeferencing has increased with year of publication. Thirty references were obtained for the period 1999–2007, and 73.3% were georeferenced compared with 15.62% of 32 references for 1998–1980. Of 27 reports that had a georeference, 59.26% were to the nearest minute and 40.7% were of greater precision. One-minute precision suggests that a gazetteer was used; greater precision suggests a GPS. Of records with a georeference, only 7.4% had information about how they obtained the georeference, 11.1% had the date of the georeference, and 3.7% reported the datum. Figure 1 also shows that the frequency of greater precision georeferences has increased with time but is still not universal. The earliest record with greater than 1-min precision was published in 1993, but the method of obtaining this



**Fig. 1.** A sample of publications reporting new mosquito distribution records and whether they include georeferences, according to year of publication. Georeferences to 1-min accuracy are indicated by closed circles, those of >1-min accuracy by open circles, and the absence of a circle indicates that no georeference was reported.

georeference was not given. Of the group to the nearest minute precision, 50.0% had voucher information, whereas this was 90.9% for the greater precision group. It seems that the attention to location detail is an indicator of detail for other aspects of the collection. Considering all records, 88.7% had the date of collection to the day, 90.3% had the method of collection, 100% reported the life stage of the collection, 66.1% kept a voucher specimen, 56.0% reported the identification method, 0% recorded the date of the identification, and 69.4% recorded information (such as the involvement of a taxonomist or appropriately detailed identification methodology) sufficient for us to ascertain the reliability of the identification.

## Discussion

The basic information needed to meaningfully map a mosquito collection event is the georeference followed by the species identification and date of capture. Other information, such as ecology and environmental conditions, may also be recorded to add value to this basic information. We have shown that the quality of the data recorded for collection events has increased since the 1980s but that improvements are still possible. Our "litmus test" for judging data quality was the completeness and quality of reports of the first occurrence of a mosquito species for locations within the North American region. We showed that these reports still require more detail regarding the datum and how the geocoordinates were obtained. We also suggest that the precision and accuracy of geocoordinates could be improved by more frequent use of such tools as a GPS.

Our study suggests that attention to detail as regards georeferencing goes hand-in-hand with the likelihood that authors will lodge voucher material. Voucher specimens including DNA extracts should be kept in cases of first occurrence records to allow both confirmation of identification and genetic analysis. Misplaced identity is one possible explanation for the apparent appearance and subsequent disappearance of a

species from an area. The discovery of cryptic species means that the identity of a reported species occurrence may need to be reassessed, possibly by use of morphological and genetic markers. In addition, the origin and mode of transport of an introduced mosquito species may be able to be ascertained if genetic material is available. For example, forensic studies of the introduction of *An. arabiensis* Patton to Brazil in the 1930s have recently been possible because voucher specimens were located that could be used for genetic analysis (Parmakelis et al. 2008). The frequency at which vouchers were kept in the studies considered here was much higher than that reported by Bortolus (2008) for ecological papers but could still be improved.

In the following, we review some basic concepts regarding georeferencing and how they might apply to mosquito species collection records.

**Basic Concepts for Georeferencing Mosquito Collection Data.** Two important concepts for georeferencing are accuracy and precision. Accuracy is a measure of how well data represent true values, whereas precision describes the finest unit of measurement used to express that value. Thus, knowing only that a mosquito was caught somewhere in Africa, we could map this collecting event with high precision as a point (e.g., the centroid of Africa in decimal degrees to 10 decimal places), but this point would not be very accurate. At the equator, records of locations to the nearest degree precision have an accuracy of  $\approx 157$  km, the nearest minute of 2.6 km, and the nearest second of 44 m (Chapman and Wieczorek 2006). Geodatabases such as Mosquitomap use decimal degree coordinates but simply converting a 1-min precision (1/60th of a degree) georeference to decimal degrees to two decimal places (1.57 km at the equator) would only introduce false precision. This is why it is important to record the original (verbatim) longitude and latitude to allow error estimation. Even if another geographic coordinate system is used, such as the Universal Transverse Mercator or Military Grid Reference System, the original data should be recorded. A GPS should be set to read in decimal degrees and a high precision to avoid errors in the conversion from another coordinate system. A measurement in decimal degrees given to five decimal places is more precise than a measurement to the nearest second, and more precise than a measurement in degrees decimal minutes given to three decimal places (Chapman and Wieczorek 2006).

The value in having an error estimate for a mosquito collection georeference is that it allows the database user to determine how accurate the georeference is, affecting such decisions as to what spatial resolution of remotely sensed data are appropriate for analyses of mosquito occurrence data. GPS receivers often have a function to determine the estimated accuracy of a given reading, and this ideally should be reported with the collection details. The point radius method (Wieczorek et al. 2004) portrays error as a radius around a geocoordinate; in the case of the African example above, the resulting circle would be very large indeed. Various sources of error contribute

to the radius of the error around a georeference (Chapman and Wieczorek 2006).

We have alluded above to the precision of the georeference as one source of error in a posteriori estimates. Failure to account for the datum is another error source. A datum is a geodetic reference system that specifies the size and shape of the earth, and the base point from which the latitude and longitude of all other points on the earth's surface are referenced. Many datums are available for different parts of the world but most spatial databases use the World Geodetic System 1984 datum (WGS84). Geographic translation software, such as Geotrans (<http://gcmd.nasa.gov/records/GEOTRANS.html>), can be used to convert geocoordinates in other datums to WGS84, and the input information should be recorded in the database (e.g., Geotrans 2.2.6: from "Indian 1954, Thailand, Bangkok" Everest to WGS84 using 47PQS6416, D. Foley Apr2008). Thus, a location in Alaska at  $-164.75423^{\circ}$  longitude,  $60.50356^{\circ}$  latitude (North American datum of 1927) would translate to  $-164.75662^{\circ}$  longitude,  $60.50273^{\circ}$  latitude (WGS84 datum). Not knowing the datum could add an error of up to 3.552 km to a georeference (Wieczorek et al. 2004). Accounting for and recording the datum, by making sure that a GPS unit is set to WGS84, reduces unnecessary error in the calculation of the spatial accuracy of a collection location.

Many mosquito collection records are georeferenced only according to the nearest town rather than the actual collection site, and this is especially problematic for a posteriori georeferencing if the town has a common name and there is no other information (e.g., populated places called Midway, appear 24 times in the state of Texas, according to Biogeomancer 1.2.1). Even when more information is available, such as an offset (e.g., 5 km N of Midway, TX), one must consider how big Midway is (from the center to the furthest extremity, i.e., the "extent"), and whether 0 km from Midway is at the center of the extent, the main post office, the courthouse, or at the edge of Midway. These problems are compounded if Midway has grown considerably in the years since the collection was undertaken. In addition, the error inherent in the precision of the distance offset (e.g.,  $5 \pm 1$  km from Midway) and the direction (e.g.,  $N \pm 45^{\circ}$  of Midway) must be estimated. Finally, was the distance calculated by road or by air? Errors such as these are usually cumulative, thereby increasing the radius of uncertainty, and reducing the value of the record.

It is important to record the source of the georeference and other details such as datum, scale, and coordinate precision that will enable the calculation of spatial error. Common sources for georeferences are maps, gazetteers, GPS, and online mapping tools such as Google Earth. For maps, the name (e.g., Series L509 Sheet NE 47-8), scale (e.g., 1:250,000), and datum (e.g., Grid IVB) are important information. Joint Operation Graphics and U.S. Geological Survey maps are more accurate than most other maps; therefore, this information should be noted. For gazetteers, the name, version, and date accessed should be recorded

(e.g., Gazetteer; Biogeomancer 1.2.1, February 2009). The make, model, reported accuracy, and datum of a GPS unit should be recorded (e.g., GPS; Magellan 315; 10 m accuracy; WCS84). Google Earth does not publish accuracy figures for its satellite imagery, and guidelines for accessing accuracy have not been published. However, it is likely that the accuracy would be similar to GPS. The name, version, and date accessed of any online mapping tools should be recorded (e.g., Google Earth, version 4.2, February 2009).

This information can be used to estimate spatial accuracy through the MaNIS/HerpNet/ORNIS Georeferencing calculator (<http://manisnet.org/gci2.html>). Recording the input parameters in addition to the results of the calculation allows others to validate the assumptions that went into the calculation (e.g., MaNIS Georef. Calc. Named place; Gazetteer; decimal deg.; WGS84; nearest minute precision; 5 km extent; using 3.88488 11.45022, D. Foley, November 2008). Biogeomancer automatically calculates spatial accuracy of gazetteer data (including any offsets) but is less transparent about input parameters. In this case, the user should record the complete results including the version of Biogeomancer, and the gazetteer name, and the input text, feature code, and feature category, to allow others to assess the results [e.g., Biogeomancer 1.2.1. 160 km W of Altamira GeoNet Names (NIMA):64737432: populated places); Pará (GADM:682:countries, first order divisions); Brazil (GADM:32:countries) ].

**Recommendations for Recording Mosquito Collection Data.** Given the challenges to accurate georeferencing discussed above, what advice can be given to entomologists who collect mosquitoes? Among the recommendations to collectors in Chapman and Wieczorek (2006) are (1) make sure the datum is recorded with all GPS readings, as well as the accuracy reported by the GPS, and the make of the GPS receiver used; (2) be consistent in the use of a standard coordinate system (e.g., use decimal degrees wherever possible); and (3) allow validation of geocoordinates by including a description of the locality in a clear and consistent manner, e.g., include nearest named place and offsets rather than vague terms such as "near," and record "by road" or "by air." The accuracy of GPS geocoordinates can be improved by averaging the results of multiple observations at a single location (McElroy et al. 1998). Canopy cover can interfere with the satellite signal but if one location can be georeferenced by a GPS, nearby locations can be georeferenced using distance and direction offsets from the clear location. Interestingly, a GPS may not be very accurate for recording elevation. Chapman and Wieczorek (2006) noted that "... the height displayed by a GPS receiver is actually the height in relation to an ellipsoid as a model of the Earth's surface, and not a height based on mean sea level, or to a standard height datum. . ." Examples of good and bad locality descriptions are given at [http://mvz.berkeley.edu/Locality\\_Field\\_Recording\\_examples.html](http://mvz.berkeley.edu/Locality_Field_Recording_examples.html). A modified example of a good locality description from this website is:

**Table 1.** Selection of data fields from Mosquitomap (<http://www.mosquitomap.org/>) showing parent and child (indented) controlled vocabulary terms

Collector	Subterranean
GlobalUniqueIdentifier	Polluted water
GenBankNumber	Exposed pool, puddle
BasisOfRecord	Exposed pond, borrow pit
Source	Exposed stream, ditch, channel
RelatedInformation	Forest pool
InformationWithheld	Forest pond
Genus	Forest stream
SubGenus	Gravel stream bed
Species	Rock-pool
ScientificName	Salt-water pool
AuthorYearOfScientificName	Salt-water pond
IdentifiedBy	Salt-water marsh
DateIdentified	Ricefields, flooded field
IdentificationMethod	Marsh
EarliestDateCollected	Swamp
LatestDateCollected	Lake
VerbatimDateOfCollection	Other = [name]
TimeOfCapture	LarvalHabitatCondition
Country	DistanceToHouseInMeters
StateProvince	DegreeOfShade
County	DistanceAboveSurfaceInMeters
Locality	VerbatimElevation
DecimalLongitude	AssociatedParasite
DecimalLatitude	LifeStageOfParasite
GeodeticDatum	IdentificationMethodForParasite
VerbatimLongitude	NumberOfMosquitoesTestedForParasite
VerbatimLatitude	RemarksAboutParasite
VerbatimCoordinates	
VerbatimCoordinateSystem	
CoordinateUncertaintyInMeters	
GeoreferenceProtocol	
GeoreferenceSources	
GeoreferenceRemarks	
IndividualCount	
Sex	
LifeStage	
Remarks	
CollectionMethod	
From colony	
Egg collection	
Egg, Larva collection ovitraps	
Larval collection	
Pupal collection	
Adult, emergence traps	
Adult, swarming collections	
Adult, animal landing/biting	
Adult, human landing/biting	
Adult, resting outdoor	
Adult, animal shed resting	
Adult, house resting	
Adult, with non-attraction traps	
Adult, with attraction traps	
Other = [name]	
CollectingEffortInHours	
LarvalHabitatType	
Bamboo	
Pitcher plant	
Fallen fruit husk	
Fallen leaf	
Tree-hole	
Tree rot-hole	
Leaf axil	
Snail shell	
Crab-hole	
Animal foot-print	
Can, bottle, tyre	
Domestic water-storage	
Water tank, cistern	
Latrine, septic tank	
Well	

**Table 1. Continued**

Subterranean
Polluted water
Exposed pool, puddle
Exposed pond, borrow pit
Exposed stream, ditch, channel
Forest pool
Forest pond
Forest stream
Gravel stream bed
Rock-pool
Salt-water pool
Salt-water pond
Salt-water marsh
Ricefields, flooded field
Marsh
Swamp
Lake
Other = [name]
LarvalHabitatCondition
DistanceToHouseInMeters
DegreeOfShade
DistanceAboveSurfaceInMeters
VerbatimElevation
AssociatedParasite
LifeStageOfParasite
IdentificationMethodForParasite
NumberOfMosquitoesTestedForParasite
RemarksAboutParasite

Modoc National Wildlife Refuge, 2.8 miles S and 1.2 miles E junction of Hwy. 299 and Hwy. 395 in Alturas, Modoc Co., CA.

Lat/Long/Datum: 41.45063, -120.50763 (WGS84)

Elevation: 1,330 ft; GPS accuracy: 24 ft; extent: 150 ft

References: Garmin Etrex Summit GPS for coordinates and accuracy, barometric altimeter for elevation

A paper or digital record should be maintained as much more information can be recorded than on a label affixed to a specimen pin or on a label associated with a slide-mounted specimen. A field collection form can be downloaded from <http://www.wrbu.org/Techniques.html>, and methodology for printing labels with barcode information can be found at <http://www.discoverlife.org/label/>.

Our experience entering collection records into the Mosquitomap geodatabase has convinced us of the need for mosquito collectors to avoid vague terminology and to adopt standards when reporting their data. The use of controlled vocabulary terms would greatly assist the compilation and searching of collection data. Mosquitomap has >60 categories of information, which represents a trade-off between specificity and utility. Many more categories could be constructed but this increases the workload of data recorders and diminishes their ability to summarize collection information. Table 1 shows a selection of the data fields currently in Mosquitomap, some of which use controlled vocabulary entries. For instance, "LarvalHabitatType" presently comprises 33 controlled vocabulary types and "CollectionMethod" 15 types (see <http://www.mosquitomap.org/> for more details). Standardization of such information is notoriously difficult as categories overlap, are not always clearly defined and vary from place to place (see Laird 1988

for a review of competing larval habitat schemas). Fitting previously recorded information into the Mosquitomap schema can be a challenge, for example, it is not clear how "marsh" is different from "swamp," and how best to classify habitat types that use the adjectives "swampy" and "marshy" in their description. To avoid omitting valuable information, original (verbatim) habitat descriptions and other information are reproduced in the "Remarks" field of Mosquitomap. Although, the Darwin Core schema is used as a basis for Mosquitomap, additional categories have been devised especially for mosquito collection data. These data include information about associated parasites, degree of shade, collecting effort, and distance to houses (Table 1). If mosquito workers are cognizant of the variety and form of collection information recorded in online databases such as Mosquitomap, published reports would be more complete, and data more easily compared between collections and collectors. We have established an online forum ([www.wrbu.org/forums/index.php](http://www.wrbu.org/forums/index.php)) to maximize the involvement of the mosquito community in the development of collection data standards and of Mosquitomap.

We recommend that Journals include in their "Advice to contributors" minimum standards for reporting collection data, at least for first records of a mosquito species for an area. Georeferencing best practice should be encouraged, including the use of GPS or other high accuracy georeferencing tools. Minimum reporting should include date of collection, collector's name, method of collection, species identification, identifier's name, date of identification, method of identification, whether it is an observation or preserved specimen, voucher information, number and life stage of specimens, and collection location (geocoordinates to at least five decimal places, datum, method of obtaining geocoordinates, make and model of GPS, and location description including offsets). The use of a controlled vocabulary keyword such as "collection records" would assist literature searches for mosquito collection data. In addition, a global unique identifier (GUID) can be assigned to a collection record, which could be published, similar to a GenBank/EMBL accession number for a DNA sequence. In Mosquitomap, the GUID is the combination: "Institution code:collection code:catalog number." Ecological niche models and other distribution models derived from these points can list the GUIDs for data points used in the models.

Mosquitomap contains point collection data and species distribution models. We anticipate that spatially georeferenced collection data will be used for a variety of purposes, including studies that relate locations to remotely sensed data. Mosquito geocoordinates should only be used with satellite data whose age and resolution matches the dates and spatial accuracy of collection data. Data users should also be aware of the potential biases within collection data

(Foley et al. 2008). Despite these caveats, mosquito collection data are a valuable resource that are often costly to obtain or may be from areas that are no longer easy to sample. We think that the proposals listed here will increase the value of collection data for entomologists and the wider scientific community.

### Acknowledgments

We thank J. Wieczorek for useful discussions involving the point radius method. This study was made possible with funding from the U.S. Department of Defense Global Emerging Infections Surveillance and Response System, Silver Spring, MD (Project G00018 08 WR), and the Global Biodiversity Information Facility (Project 2007-73).

### References Cited

Belkin, J. N., C. L. Hogue, P. Galindo, T.H.G. Aitken, R. X. Schick, and W. A. Powder. 1965. Mosquito studies (Diptera, Culicidae). II. Methods for the collection, rearing and preservation of mosquitoes. Contrib. Am. Entomol. Inst. 1: 19–78.

Bortolus, A. 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio* 37: 114–118.

Chapman, A. D., and J. Wieczorek (eds.). 2006. Guide to best practices for georeferencing. Global Biodiversity Information Facility, Copenhagen, Denmark.

Darsie, R. F., Jr., and R. A. Ward. 2000. Summary of new distribution records for mosquito species in the United States and Canada for the period 1981–99. *J. Am. Mosq. Control Assoc.* 16: 1–4.

Faran, M. E., C. Burnett, J. J. Crockett, and W. L. Lawson. 1984. A computerized mosquito information and collection management system for systematic research and medical entomology (Diptera: Culicidae). *Mosq. Syst.* 16: 289–307.

Foley, D. H., A. L. Weitzman, S. E. Miller, M. E. Faran, L. M. Rueda, and R. C. Wilkerson. 2008. The value of georeferenced collection records for predicting patterns of mosquito species richness and endemism in the Neotropics. *Ecol. Entomol.* 33: 12–23.

Laird, M. 1988. The natural history of larval mosquito habitats. Academic, London, United Kingdom.

McElroy, S., I. Robins, G. Jones, and D. Kinlyside. 1998. Exploring GPS, a GPS users guide; the Global Positioning System Consortium (GPSCO). Global Positioning System Consortium, Bathurst, Australia.

Parmakelis, A., M. A. Russello, A. Caccone, C. B. Marcondes, J. Costa, O. P. Forattini, M.A.M. Sallum, R. C. Wilkerson, and J. R. Powell. 2008. Historical analysis of a near disaster: *Anopheles gambiae* in Brazil. *Am. J. Trop. Med. Hyg.* 78: 176–178.

White, K. E., and G. Grodhaus. 1972. Computer information retrieval system for California mosquito collection records. *Calif. Vector News* 19: 27–39.

Wieczorek, J., Q. Guo, and R. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int. J. Geogr. Inf. Sci.* 18: 745–767.

Received 5 December 2008; accepted 26 February 2009.